Coffee Shop Sales Data Analysis

```
!pip install pandas openpyxl matplotlib seaborn plotly
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from google.colab import files
```

Step-1: Upload and Load the Data

```
file_name = "Coffee Shop Sales.xlsx"
df = pd.read_excel(file_name)
df.head()
```

	transaction_id	transaction_date	transaction_time	transaction_qty	store_id	store_location	product_id	unit_price	product_ca
0	1	2023-01-01	07:06:11	2	5	Lower Manhattan	32	3.0	
1	2	2023-01-01	07:08:56	2	5	Lower Manhattan	57	3.1	
2	3	2023-01-01	07:14:04	2	5	Lower Manhattan	59	4.5	Drinking Ch
3	4	2023-01-01	07:20:24	1	5	Lower Manhattan	22	2.0	
4	5	2023-01-01	07:22:41	2	5	Lower Manhattan	57	3.1	

Step-2: Explore and Clean the Data

```
df.info()
df['transaction_date'] = pd.to_datetime(df['transaction_date'])
df['Total Sales'] = df['transaction_qty'] * df['unit_price']
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149116 entries, 0 to 149115
Data columns (total 12 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   transaction_id    149116 non-null  int64
 1   transaction_date  149116 non-null  datetime64[ns]
 2   transaction_time  149116 non-null  object
 3   transaction_qty   149116 non-null  int64
 4   store_id          149116 non-null  int64
 5   store_location    149116 non-null  object
 6   product_id        149116 non-null  int64
 7   unit_price        149116 non-null  float64
 8   product_category  149116 non-null  object
 9   product_type      149116 non-null  object
 10  product_detail    149116 non-null  object
 11  Total Sales       149116 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(4), object(5)
memory usage: 13.7+ MB
```

Step-3: Perform Data Analysis

i) Total Sales by Product Category: Analyze total sales based on product_category.

```
category_sales = df.groupby('product_category')['Total Sales'].sum().sort_values(ascending=Fals
print(category_sales)
```

```
product_category
Coffee                269952.45
Tea                   196405.95
Bakery                 82315.64
Drinking Chocolate     72416.00
Coffee beans           40085.25
Branded                13607.00
Loose Tea              11213.60
Flavours                8408.80
Packaged Chocolate      4407.64
Name: Total Sales, dtype: float64
```

ii) Total Sales by Store Location: Analyze sales based on store_location.

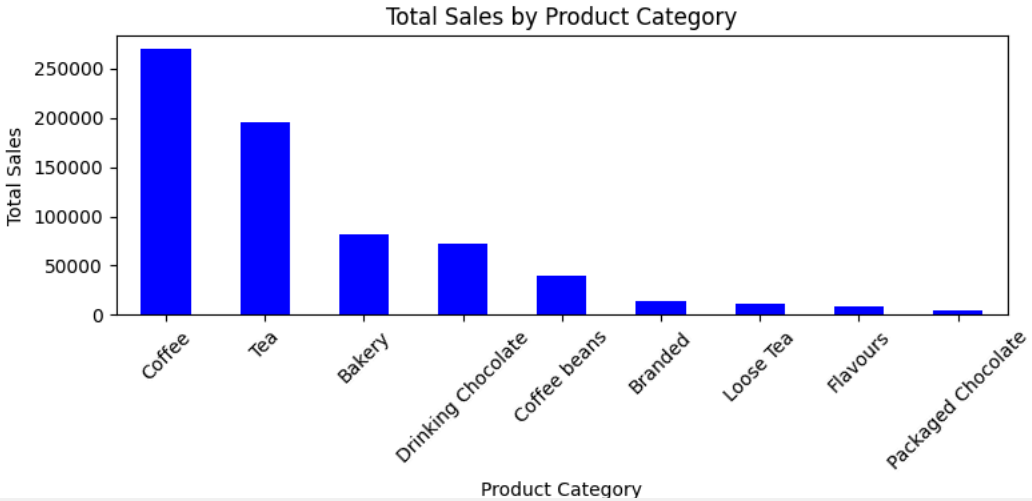
```
location_sales = df.groupby('store_location')['Total Sales'].sum().sort_values(ascending=False)
print(location_sales)
```

```
store_location
Hell's Kitchen     236511.17
Astoria            232243.91
Lower Manhattan    230057.25
Name: Total Sales, dtype: float64
```

Step-4: Matplotlib Visualizations

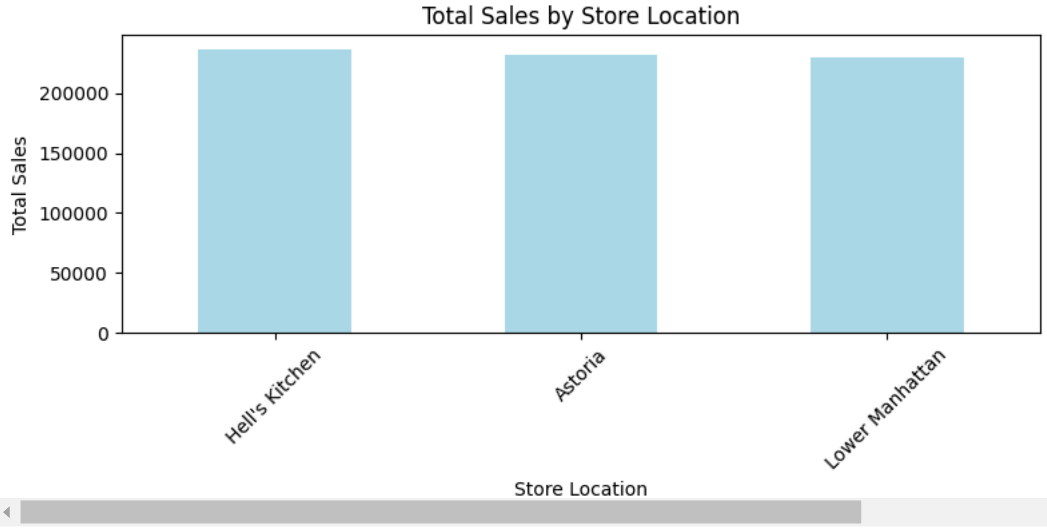
i) Bar Plot for Sales by Product Category

```
plt.figure(figsize=(8,4))
category_sales.plot(kind='bar', color='blue')
plt.title('Total Sales by Product Category')
plt.xlabel('Product Category')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



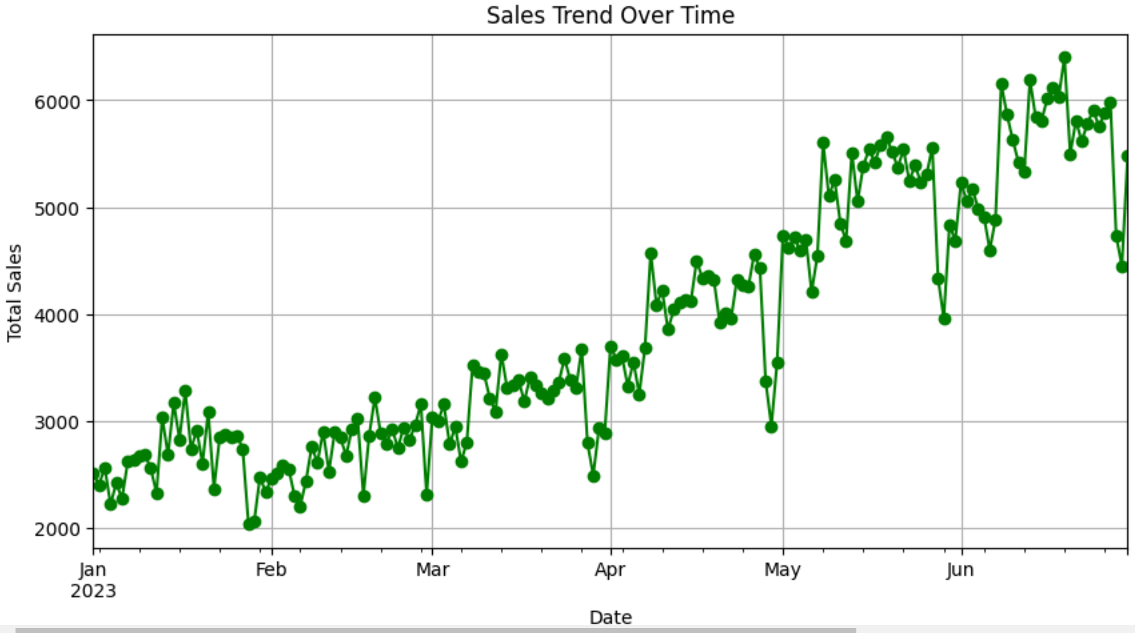
ii) Bar Plot for Sales by Store Location

```
plt.figure(figsize=(8,4))
location_sales.plot(kind='bar', color='lightblue')
plt.title('Total Sales by Store Location')
plt.xlabel('Store Location')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



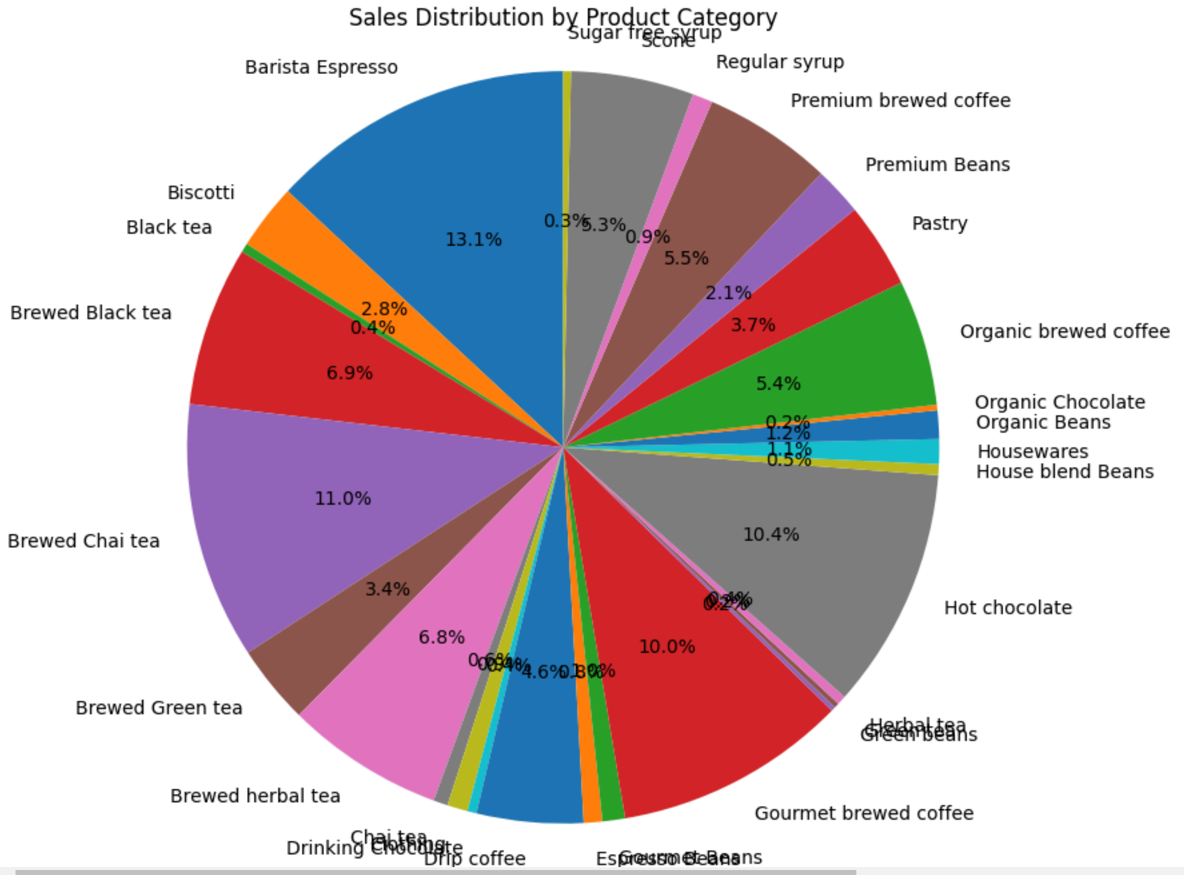
iii) Line Plot for Sales Over Time

```
daily_sales = df.groupby(df['transaction_date'].dt.to_period('D'))['Total Sales'].sum()
plt.figure(figsize=(10,5))
daily_sales.plot(kind='line', marker='o', color='green')
plt.title('Sales Trend Over Time')
plt.xlabel('Date')
plt.ylabel('Total Sales')
plt.grid(True)
plt.show()
```



Step-5: Pie Chart for Sales Distribution by Product Type

```
product_type_sales = df.groupby('product_type')['Total Sales'].sum()
plt.figure(figsize=(8,8))
plt.pie(product_type_sales, labels=product_type_sales.index,autopct='%1.1f%%',startangle=90)
plt.title('Sales Distribution by Product Category')
plt.axis('equal')
plt.show()
```



SUMMARY:

This project analyzed coffee shop sales data, providing insights into total sales by product category, store location, and sales trends over time. Visualizations such as bar charts, line plots, and pie charts were created using Matplotlib to help understand the sales distribution and trends effectively.

Thank you for taking the time to review this project.

Created by: ***Unnati Awaghad***